MIRROS: Phase 3 trial with time-to-event endpoint, a cure proportion, and a futility interim analysis using response

Kaspar Rufibach  
Methods, Collaboration & Outreach Group, Department of Biostatistics, Roche Basel  
Joint work with Dominik Heinzmann and Annabelle Monnet (Roche, Basel)  
SMi Adaptive Designs in Clinical Trials Conference London, 2\textsuperscript{nd} April 2019
Agenda

1. Acute Myeloid Leukemia
2. Clinical development plan
3. Key questions of MIRROS
4. Implementation features
5. Health authority feedback
6. Conclusions
What MIRROS is NOT:

An adaptive trial.

A seamless phase 2/3 trial.
What MIRROS IS:

A phase 3 trial with a futility interim.
Acute Myeloid Leukemia
Acute Myeloid Leukemia

Rare malignant blood disease.

Most common leukemia, lowest survival rate in adults: median survival $\leq 1y$.

Recurrent life-threatening infections.

Chemotherapy: modest benefit without cure.

Stem cell transplant:

- “Bridge-to-transplant”: Goal of any therapy. Needs complete response (CR) to initial therapy.
- Only way to survive AML.
Standard of care

No standard regimen for relapsed or refractory (R/R) AML. Breems et al. (2005)

No new drug approved for treatment of AML in over 50 years! Bose et al. (2017)

THIS is unmet medical need!
**Idasanutlin**

**p53**: Tumor suppressor, many mechanisms of anticancer function.

**Mouse double minute 2 homolog (MDM2)**: Negative regulator of p53 tumor suppressor.

**Idasanutlin**: binds to MDM2 ⇒ prevents p53 - MDM2 interaction ⇒ (re-)activation of p53 ⇒ **reinstalls anti-tumor capacity of p53**.
Clinical development plan
Clinical development plan for Idasanutlin

Need for acceleration:

- Very high unmet medical need in R/R AML.
- Early phase results with Idasanutlin encouraging.
- Competitive landscape and economic constraints: Lean program only way to receive internal approval for pivotal trial.
- Willingness to trade-off risk reduction from randomized P2 against increased speed.
Skip or integrate Phase 2?

Assume we have **successful P1**.

Purpose of futility interim: optimize $P(\text{stopping @ interim} \mid H_0)$.

Hunsberger et al. (2009):
- **Integrate** P2 into P3: futility interim based on intermediate endpoint.
- **Skip** P2: futility interim based on P3 primary endpoint.

If trial
- stops at futility interim: basically performed randomized P2.
- passes futility interim: P3 pivotal trial well on its way.

Key advantage of setup: Decision to proceed to full P3 part based on randomized comparison. **Parmar et al. (2008)**
MDM2 Idasanutlin in Relapsed Refractory AML for OS.

- **Population:** R/R AML.
- **Comparison:** Idasanutlin + cytarabine vs. placebo + cytarabine.
- **Phase III, 2:1 randomized, double-blind, placebo-controlled** clinical trial.
- **Primary endpoint:** overall survival.
- **Planned recruitment:** 374 patients (wild-type sample, + 66 mutant patients).

https://clinicaltrials.gov/ct2/show/NCT02545283
Key questions of MIRROS
Key questions of MIRROS

1. Primary endpoint OS. Sample size with \textit{cure proportion} in both arms?
2. Base \textit{interim} on OS or something else? If the latter, what?
3. How to compute \textit{operating characteristics} of interim analysis?
Cure proportion model

See e.g. Sun et al. (2018).

Let

- $S_i^*$, $f_i^*$: survival and density functions of uncured patients.
- $p_i$: proportions of patients cured.

Survival and hazard function in each treatment arm ($t \geq 0$):

$$S_i(t) = p_i + (1 - p_i)S_i^*(t),$$

$$h_i(t) = \frac{(1 - p_i)f_i^*(t)}{p_i + (1 - p_i)S_i^*(t)}.$$

Ratio of hazard functions:

$$\theta(t) = \frac{h_2(t)}{h_1(t)} = \frac{(1 - p_2)f_2^*(t)}{f_1^*(t)} \frac{p_1 + (1 - p_1)S_1^*(t)}{p_2 + (1 - p_2)S_2^*(t)}.$$

Even if both $S_i^*$ exponential $\Rightarrow \theta(t)$ depends on time (if $\geq 1 p_i$ is $> 0$).
Cure proportion model – assumptions

What if we ignored cure proportions and simply computed necessary events $d$ using Schoenfeld’s formula?

- Study will (typically) be underpowered.
- Time to clinical cutoff will be underestimated.

Control arm, based on historical data, $H_0$:

- Median OS 6m.
- Cure: 0.080.

Targeted effect size treatment arm (for 85% power, $H_1$):

- Median OS 9m.
- Cure: 0.161 (see later for justification).
Cure proportion model – assumptions

**Survival functions**
- Probability to survive over overall survival (months)
- Control and Treatment curves

**Hazard functions**
- Hazard function over overall survival (months)
- Control and Treatment curves
Cure proportion model – assumptions

Survival functions

- Control
- Treatment
- Cure proportions

Overall survival (months)

Probability to survive

Ratio of hazard functions

Hazard ratio = 2/3 and 1

Overall survival (months)
Cure proportion model – sample size

To find sample size:

- Compute necessary events \( d_0 \) using Schoenfeld’s formula.
- **Simulate** from assumed \( S_i \)'s, compute power for grid of \( d = d_0, \ldots, d_1 \).
- Choose \( d \) such that (unweighted) logrank test gives targeted power.

**MIRROS**: 2-sided \( \alpha = 0.05, \beta = 0.15 \), some accrual and drop-out assumption.

<table>
<thead>
<tr>
<th>Assumption</th>
<th>( S_1^{-1}(0.5) )</th>
<th>( S_2^{-1}(0.5) )</th>
<th>( p_1 )</th>
<th>( p_2 )</th>
<th>( d )</th>
<th>power</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIRROS</td>
<td>6.0</td>
<td>9.0</td>
<td>0.080</td>
<td>0.161</td>
<td>275</td>
<td>0.852</td>
<td>38.8</td>
</tr>
<tr>
<td>PH, no cure</td>
<td>6.0</td>
<td>9.0</td>
<td>0</td>
<td>0</td>
<td>246</td>
<td>0.858</td>
<td>29.2</td>
</tr>
<tr>
<td>MIRROS with #events for PH, no cure</td>
<td>6.0</td>
<td>9.0</td>
<td>0.080</td>
<td>0.161</td>
<td>246</td>
<td><strong>0.810</strong></td>
<td>33.7</td>
</tr>
</tbody>
</table>
Cure proportion model – effect quantification

Cure proportion model – no proportional hazards. Unweighted logrank...
- ...not most powerful test, but loss modest (see above).
- ...still valid test, i.e. protects type I error.

How to quantify effect?
- Kaplan-Meier estimates provide entire information in data.
- Desire to summarize effect in one number.
- Hazard ratio and logrank test: if NPH, estimand depends on censoring distribution!
- Regulatory environment: typically accepted to reject $H_0$ using valid test, and then quantify effect differently.
- MIRROS: violation of PH only very late. Give hazard ratio, and estimate of cure proportion difference.

Rufibach (2019) has extended discussion in estimand context.
Futility interim analysis

Mitigate risk if drug does not work (sufficiently).

Planned after 120 patients are recruited.

Why not use OS for interim decision?

- 53 (under $H_0$) and 46 deaths (under $H_1$) expected at interim. Substantial uncertainty.
- Cures have not happened yet at the interim.
- Confounding by early (mainly safety-related) deaths.

Bottom line: interim is too early for OS to be meaningful endpoint.
Intermediate endpoint

Complete response:

- Sufficiently associated with OS.
- CR **necessary** for good OS / cure: Patient needs CR to have chance for cure, via bridge-to-transplant.
- Odds ratio as effect measure.

Futility interim is **non-binding**. Why do we need to model it at all?

- How to choose interim boundary on CR?
- Decision-makers want to be able to trade-off

\[
\text{False Positive} = P(\text{continue @ interim} \mid H_0) \\
\text{vs.}
\]

\[
\text{False Negative} = P(\text{stop @ interim} \mid H_1).
\]

If futility based on OS ⇒ conditional power.

If CR is intermediate endpoint: **mechanistic simulation model**.
Mechanistic simulation model

Control

1-p_{CR,1}

Non-responders

1-p_{L,1}

Short-term survivor

S_{N,1}(t)

Responders

p_{CR,1}

Long-term survivor

S_{S,1}(t)

Experimental

1-p_{CR,2}

Non-responders

1-p_{L,2}

Short-term survivor

S_{N,2}(t)

Responders

p_{CR,2}

Long-term survivor

S_{S,2}(t)

S_{L,1}(t)

S_{L,2}(t)
Mechanistic simulation model

Connects CR to OS.

Need to inform all assumptions:

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Control arm</th>
<th>Treatment arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival function of non-responders</td>
<td>$S_{N,1}$</td>
<td>$S_{N,2}$</td>
</tr>
<tr>
<td>Probability to have CR</td>
<td>$p_{CR,1}$</td>
<td>$p_{CR,2}$</td>
</tr>
<tr>
<td>Probability to be long-term responder</td>
<td>CR</td>
<td>$p_{L,1}$</td>
</tr>
<tr>
<td>Survival function of short-term responders</td>
<td>$S_{S,1}$</td>
<td>$S_{S,2}$</td>
</tr>
<tr>
<td>Survival function of long-term responders</td>
<td>$S_{L,1}$</td>
<td>$S_{L,2}$</td>
</tr>
<tr>
<td># patients recruited per month</td>
<td>$n_{1j}$</td>
<td>$n_{2j}$</td>
</tr>
<tr>
<td>Months of recruitment</td>
<td>$j = 1, \ldots, N$</td>
<td></td>
</tr>
<tr>
<td>Total # patients recruited</td>
<td>$n_1 = \sum_{j=1}^{N} n_{1j}$</td>
<td>$n_2 = \sum_{j=1}^{N} n_{2j}$</td>
</tr>
<tr>
<td>Drop-out rate per month</td>
<td>$\tau_1$</td>
<td>$\tau_2$</td>
</tr>
</tbody>
</table>

Align parameters such that mechanistic simulation model can reproduce sample size!

P(CR) control: 0.16. Assume OR = 2.5 to improve on this with treatment ⇒ P(CR tmt) = 0.323. P(longterm survivor) = 0.5. This gives cure proportions.
Operating characteristics of various interim boundaries

False Positive = P(continue @ interim | no effect)
False Negative = P(stop @ interim | alternative used for powering)
Operating characteristics of various interim boundaries

Sweet spot: odds ratio of 2,

- False Positive = $P(\text{continue @ interim} \mid \text{no effect}) \approx 12\%$,
- False Negative = $P(\text{stop @ interim} \mid \text{alternative assumed for powering}) \approx 30\%$.

Interim decision:

- Based on independent data monitoring committee (iDMC) recommendation, i.e. sponsor blinded,
- non-binding,
- included safety criterion (molecule class toxicity) and criteria for early deaths ⇒ OS component.
Power loss of adding futility interim

Can easily get that from simulations.

- Targeted power: 85%.
- Power taking into account futility interim: 63%!
- Illustrates risk-appetite. Futility interim somehow becomes “informal efficacy interim”.
- Do we always compute the power loss when adding futility interims? Do we increase number of events to account for it?

Who cares anyway ⇒ interim passed!
Implementation features
Implementation features

A (industry) clinical trial is **not a pre-specified static** undertaking!

- Not clear whether p53 mutant patients ($\approx 15\%$) also benefit from Idasanutlin.
  - Still included, as evidence unclear and high unmet medical need.
  - But testing too late for randomization, i.e. could not stratify for p53 status.
  - Adds uncertainty to recruitment assumptions.

- Decision-makers sceptical about interim gate based on CR only. Additionally engineered EFS criterion (not discussed here).
Implementation features

A (industry) clinical trial is not a pre-specified static undertaking!

- Biomarker development: typically in Phase 2! Recommendation on biomarker development by iDMC.

- Seamless designs in general: sponsor does not get to see data for a long time. Unease for decision-makers.

- No accrual suspension for interim ⇒ data cleaning and decision needs to come fast.
Health authority feedback
Health authority feedback

FDA:

- **Preferred randomized P2.**
- Challenged lack of stratification on p53 mutation status.
- Companion Diagnostic component with blinded P2 data ⇒ not clear how to decide on development.
- **Challenged assumptions**, asked for additional sensitivity analyses.
- Concerns of early events driving interim analysis. OS not part of futility decision, but early tox deaths are.
- US sites only opened after passing the IA.

EMA:

- Agreed to accelerated development due to high unmet need.
- PH assumption discussed, support hazard ratio as appropriate effect measure.
Conclusions
Current status of MIRROS

Interim analysis passed on 17th Sept 2017.

Final analysis cutoff projected for Q4 2019.
Conclusions

- Account for power loss and timing delay if you have cure proportions.
- Think about how to quantify effect.
- Skipping / integrating P2 into P3 allows for acceleration and risk-mitigation. If you stop at interim not much is lost in fact.
- Mechanistic simulation model allows to associate binary intermediate to time-to-event primary endpoint and explore interim analysis operating characteristics.

We have implemented this in a REAL trial!
Resources


Code to reproduce simulations and plan your own trial on github: https://github.com/numbersman77/integratePhase2.git.
BBS seminar on synthetic controls

Basel Biometric Section spring seminar: **Synthetic controls - what do we need and how far can we go?**

- Basel.
- **May 10th**, 2019, 9:00-16:00.
- Speakers from industry, Flatiron, European regulators.
- Rejoinders by regulators: Norbert Benda (BfArM), Jan Müller-Berghaus (PEI), Anja Schiel (Norwegians Medicine Agency & Chair BSWP), Kit Roes (UMC Utrecht MEB and EMA BSWP), Meinhard Kieser (University of Heidelberg).
- Panel discussion with all speakers.

http://bbs.ceb-institute.org
Thank you for your attention.
References I


References II


Backup slides.
Cure proportion model – estimation


Obvious nonparametric estimate of $p$, with $\hat{S}$ Kaplan-Meier:

- $\hat{S}(t_0)$ for some $t_0 > 0$.
- Maller and Zhou (1992): Kaplan-Meier evaluated at largest observed time, censored or event, consistently estimates $p_0$ under “sufficient follow-up” condition Tsodikov et al. (2003).
- Finite sample: likely not use latest observed time to evaluate the Kaplan-Meier estimate at. Rather trade-off bias to reduce variability of estimate.
- Choose milestone $t_0$ where clinically, cure seems very plausible.
Why two models?

We have two models:

- Cure proportion model to derive sample size,
- Mechanistic simulation model to explore interim operating characteristics.

Why?

Reasons:

- Futility interim analysis has no implication on type I error ⇒ independent of key design characteristic.
- Cure proportion model:
  - Simple,
  - depends on less assumptions than mechanistic model,
  - Robust model to plan sample size.
- Mechanistic simulation model:
  - Interim setup has potential to be changed before or while study is running. Prefer not to have these changes interfere with sample size.
  - Only used for (internal) decision-making via iDMC, no filing relevance ⇒ can “afford” more modeling.
Doing now what patients need next

R version and packages used to generate these slides:

R version: R version 4.0.0 (2020-04-24)
Base packages: stats / graphics / grDevices / utils / datasets / methods / base
Other packages: biostatKR / survival / rpact / reporttools / xtable / probSuccess / cubature / pracma / mvtnorm / dplyr / readxl

This document was generated on 2020-06-28 at 16:40:23.